Statistical Inference Course Project - Part 2

Author	Date	GitHub
Oszkar Jozsa	February, 2015	[link]

Overview

Description

Now in the second portion of the class, we're going to analyze the ToothGrowth data in the R datasets package.

- Load the ToothGrowth data and perform some basic exploratory data analyses
- Provide a basic summary of the data.
- Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
- State your conclusions and the assumptions needed for your conclusions.

Solution

Exploratory data analysis

We can find a descriptive info about the "The Effect of Vitamin C on Tooth Growth in Guinea Pigs" dataset here. It says: "The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid)."

Let's look at the content of `ToothGrowth`!

```
library(ggplot2)
library(datasets)
data(ToothGrowth)
head(ToothGrowth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
str(ToothGrowth)
```

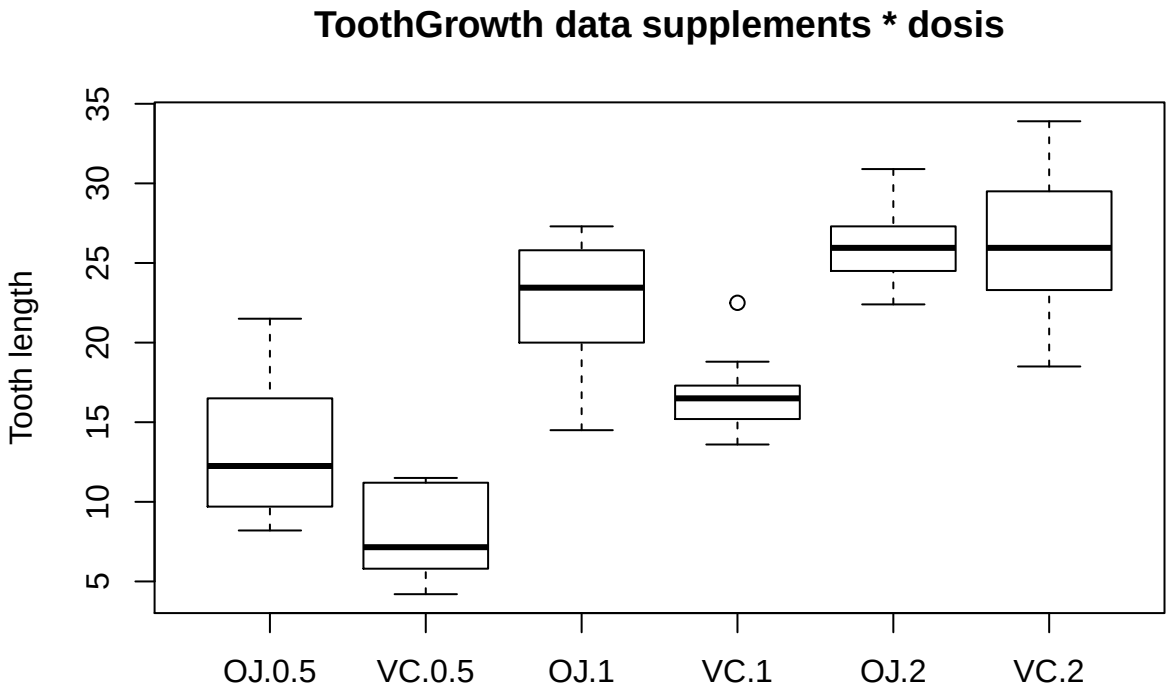
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
summary(ToothGrowth)
```

```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

So we have 2 different supplements, `VC` and `OJ` and three different doses, `0.5`, `1.0`, `2.0` (mg). These two (supplement and dosis) have effect on the length, saved in the `len` column. We can visualize the relations between these using a boxplot.

```
boxplot(len ~ supp * dose, data = ToothGrowth, ylab="Tooth length", main="ToothGrowth data supplements * dosis"
```



Now, setting aside the dosage, we'll look at the effectiveness of the two supplements.

```
OJ = c(mean(ToothGrowth[ToothGrowth$supp == "VC",]$len), sd(ToothGrowth[ToothGrowth$supp == "VC",]$len))
VC = c(mean(ToothGrowth[ToothGrowth$supp == "OJ",]$len), sd(ToothGrowth[ToothGrowth$supp == "OJ",]$len))
data.frame(OJ, VC, row.names = c("mean", "variance"))
```

```
##                 OJ        VC
## mean     16.963333 20.663333
## variance  8.266029  6.605561
```

Based on both the full box plot and on the means-variances of the two different supplements, we can conclude that the Orange Juice (`OJ`) supplement is better.

Hypothesis test

T distribution will be used for hypothesis testing.

```
t.test(len ~ supp, data = ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

We won't reject the null hypothesis as the P-value is above 0.05 and the confidence interval contains 0. This means that the effect of the two supplements fo differ.

Now test the doses (as dosage pairs):

```
t.test(len ~ dose, data = subset (ToothGrowth, dose %in% c(0.5, 1.0)))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

```
t.test(len ~ dose, data = subset (ToothGrowth, dose %in% c(0.5, 2.0)))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

```
t.test(len ~ dose, data = subset (ToothGrowth, dose %in% c(1,0, 2.0)))
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

Conclusions

The quick research of the `ToothGrowth` data showed us that both the dosage and the supplement type plays a role in tooth growth in gueina pigs. Orange juice or ascorbic acid delivery method proves to be more effective than Vitamin C. Also, bigger dosage helps growth (on the examined [0.5..2.0] interval).